microbiology Comment

Genome Update: promoter profiles

Genomes of the month

Ten new genomes have been published since last month's Genome Update. The list includes six bacterial and four yeast genomes, as listed in Table 1. A brief overview of each genome will be given below.

Acinetobacter belongs to the family *Moraxellaceae* of the *γ-Proteobacteria*, and is a neighbour of the family *Vibrionaceae*. The genome of *Acinetobacter* sp. ADP1 has been sequenced (Barbe *et al.*, 2004); this organism is a soil isolate which can be transformed easily. The genome of *Acinetobacter* sp. ADP1 is 3·6 Mbp long, with an A + T content of 60 % (see Table 1).

Bacillus thuringiensis has been used commercially for the biological control of insect pests, and this insecticidal activity is due to the ability of the organism to synthesize a large amount of protein that forms a parasporal crystal during sporulation. (Schnepf *et al.*, 1998). The *Bacillus thuringiensis* genome is 5·2 Mbp long, with an A + T content of 65 %. There are 105 tRNA genes (a high number for bacterial genomes) and the 14 rRNA genes is the highest number so far found, out of 171 sequenced bacterial genomes. It is likely that the large number of rRNA and tRNA genes are useful for high growth rates.

Two *Bartonella* genomes have been published recently (Alsmark *et al.*, 2004). *Bartonella quintana* (1·6 Mbp) and *Bartonella henselae* (1·9 Mbp) are facultative intracellular bacteria and human pathogens. *Bartonella quintana* is the causative agent of trench fever, a disease that affected more than 1 million soldiers during World War I, and it is transmitted by human body lice. It has a low coding density of 73 % compared to the other sequenced bacteria; the genome encodes a total of 1308 predicted genes (see Table 1). *Bartonella henselae* infects both humans and cats (30–60 % of domestic cats in the USA are infected with this organism). Transmission among cats is mediated by the cat flea and to humans by cat scratches or cat bite. *Bartonella henselae* has a similar coding density of 73 %, and a slightly larger genome which encodes 1612 predicted genes. The primary difference between these closely related strains is their reservoir ecology.

Erwinia carotovora is a plant-pathogenic enterobacterium and the causative agent of soft rot and blackleg potato diseases. *E. carotovora* differs from the sequenced enterobacterial human pathogens by about a third of its genome, and some of these genes are predicted to facilitate nitrogen fixation and opine catabolism. *E. carotovora* has an A + T content of 49 %, which is similar to that of *Escherichia* species, and the genome encodes a total of 4492 predicted genes, including 9 and 7 of 5S and 16S, respectively (Bell *et al.*, 2004).

Mesoplasma florum is a non-pathogenic organism and a non-motile mycoplasma species, but it is not closely related to *Mycoplasma genitalium* or *Mycoplasma pneumoniae*. It has the characteristic small size genome of *Mycoplasma* species, with only 793 kbp, and an A + T content of 73 %; there are only 29 predicted tRNAs for this slow-growing organism.

Finally, four yeast genomes have been published (Dujon *et al.*, 2004) which represent a quite diverse set of organisms within the phylum Ascomycota. The four yeast genomes shown in Table 1 are from four different branches of the ascomycete lineage. Although all four genomes contain about 6000 genes, the genome size ranges between 10 Mbp and 20 Mbp. The *Yarrowia lipolytica* genome is the largest (20·5 Mbp) and has the lowest coding density; the *Debaryomyces hansenii* genome is only 12·2 Mbp long, yet it has about 200 more genes than that of *Y. lipolytica* (see Table 1). Although the exact number of rRNAs is not known, the number of clusters is given, and Dujon *et al.* (2004) report that there are more than 105 copies of rRNA genes scattered throughout the genome of *Y. lipolytica*, in addition to the seven clusters. These four genomes, when compared to each other and other yeast genomes, shed light on possible evolutionary routes taken by each branch of the ascomycete lineage.

Method of the month – promoter profiles

By looking at the average value of DNA structural parameters for all the genes in a genome, aligned at translation start sites, it is possible to construct a 'promoter profile'. There is a general trend for the upstream regions to be AT-rich, more rigid and more easily melted than on average for the whole chromosome (Pedersen *et al.*, 2000). Different genomes can have different promoter profiles,

***Microbiology* Comment** provides a platform for readers of *Microbiology* to communicate their personal observations and opinions in a more informal way than through the submission of papers.

Most of us feel, from time to time, that other authors have not acknowledged the work of our own or other groups or have omitted to interpret important aspects of their own data. Perhaps we have observations that, although not sufficient to merit a full paper, add a further dimension to one published by others, or we may have a useful piece of methodology that we would like to share.

Guidelines on how to submit a *Microbiology* Comment article can be found in the Instructions for Authors at http://mic.sgmjournals.org

It should be noted that the Editors of *Microbiology* do not necessarily agree with the views expressed in *Microbiology* Comment.

Chris Thomas, Editor-in-Chief

Table 1. Summary of the published genomes discussed in this Update

Note that the accession number for each chromosome is the same for GenBank, EMBL and the DDBJ.

Name	Strain	Length (bp)	AT content (%)	No. of genes	tRNAs	rRNAs*	Accession no.
Acinetobacter sp.	ADP1†	3 598 621	59·6	3 325	76	7	CR543861
Bacillus thuringiensis	97-27	5 237 682	64·6	5 117	105	14	AE017355
Bartonella henselae	Houston-1	1 931 047	61·8	1 612	43	4	BX897699
Bartonella quintana	Toulouse	1 581 384	61·2	1 308	42	4	BX897700
Erwinia carotovora	SCRI1043	5 064 019	49·0	4 492	76	7	BX950851
Mesoplasma florum	L1	793 224	73·0	683	29	2	AE017263
Candida glabrata	CBS 138	12 280 357	61·4	5 272	207	2 clusters	CR380947–CR380954
Debaryomyces hansenii	CBS 767	12 220 823	61·6	6 896	205	3 clusters	CR382133–CR382139
Kluyveromyces lactis	NRRL Y-1140	10 689 156	61·2	5 331	162	1 cluster	CR382121–CR382126
Yarrowia lipolytica	CLIB99	20 502 981	50·9	6 666	510	7 clusters	CR382127–CR382132

*The rRNA column refers to the number of operons; 'clusteres' indicates the number of clusters of tandem repeated rRNA operons reported.
†Based on information from Tatiana Tatusov, from GenBank (see also http://www.genoscope.cns.fr/externe/English/Projets/Projet_DY/DY.html).

reflecting various environmental and other aspects of genome organization. Fig. 1 shows the promoter profiles for four of this month's genomes. For all the genomes, there is a strong difference between the first half of the plot (upstream from translation start sites) and the second half (downstream). As discussed in a previous Genome

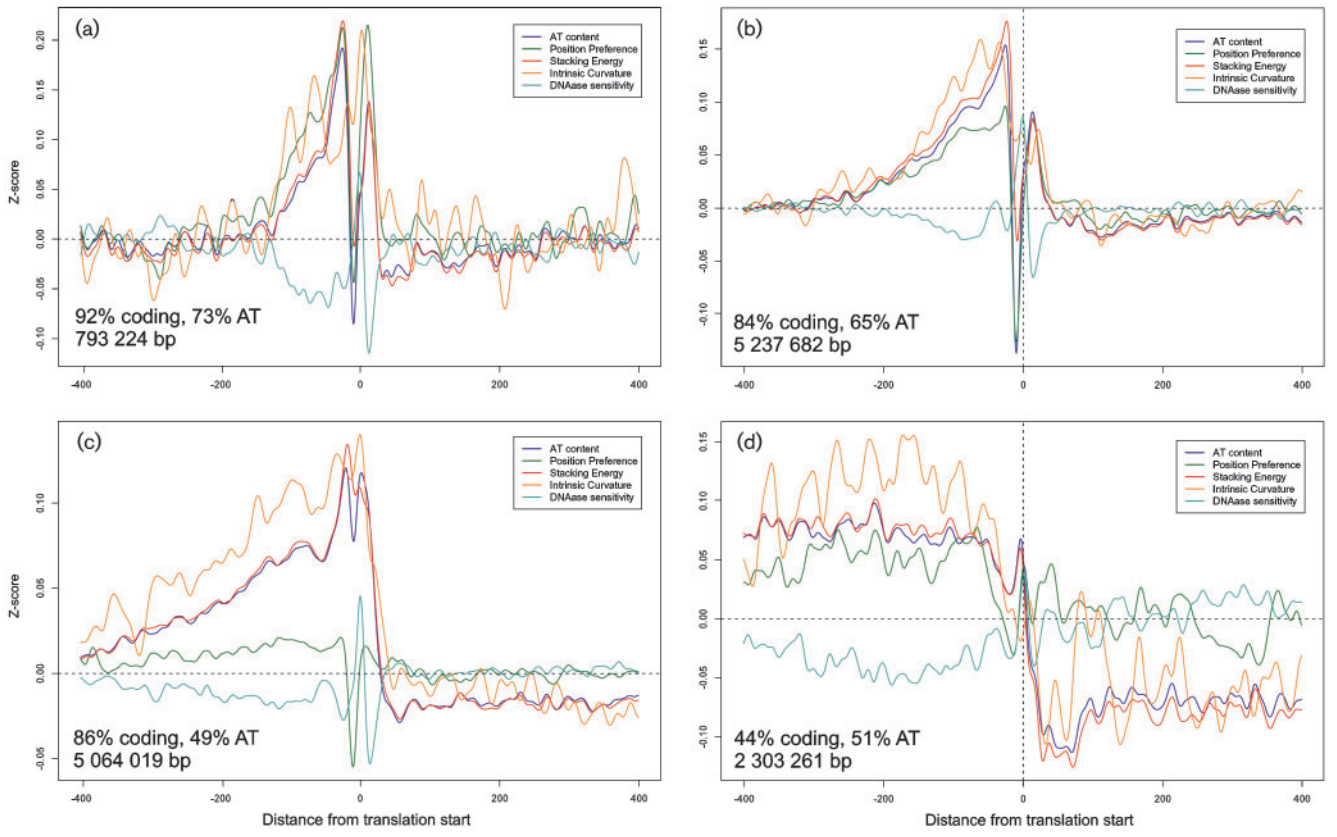

Fig. 1. Promoter profiles for the genomes of the bacteria *Mesoplasma florum* strain L1 (a), *Bacillus thuringiensis* strain 97-27 (b) and *Erwinia carotovora* strain SCRI1043 (c); also shown is the promoter profile for chromosome 1 of the yeast *Yarrowia lipolytica* strain CLIB99 (d). Each gene from the chromosome was aligned at the translation start site, and the average DNA structural property was calculated for each position in the alignment. Differences from the chromosomal average are plotted in the figures.

Update (Ussery & Hallin, 2004), for nearly all bacterial genomes sequenced, the region right before translation start is more AT-rich than the coding sequences. In the three bacterial genomes shown in Fig. 1, there is a general increase in deviation from the chromosomal average for most of the structural parameters up to the translation start site ('0' on these plots), whilst downstream the signal intensities are close to average for the whole genome. This general trend is true for most bacterial genomes, although for eukaryotes the downstream region can vary significantly from the average, as can be seen for the *Y. lipolytica* chromosome 1 plot in Fig. 1. This could be due to the much lower coding density (less than 50 %) compared with the bacterial genomes (typically around 85 % or larger). Along these lines, it is interesting to note that the *Mesoplasma florum* genome has the highest coding density of the genomes in Table 1, and it also has a more narrow range of high AT content upstream, probably reflecting a shorter distance between transcription +1 and the translation start sites. In summary, promoter profiles can be used to extract useful information about the structural characteristics of promoters from various genomes.

Supplemental web pages

Web pages containing material related to this article can be accessed from the following url: http://www.cbs.dtu.dk/services/GenomeAtlas/suppl/GenUp008/

Acknowledgements

This work was supported by a grant from the Danish National Research Foundation.

David W. Ussery, Nikolaj Tindbæk and Peter F. Hallin

Center for Biological Sequence Analysis, Department of Biotechnology, Building 208, The Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark

Correspondence: David W. Ussery (dave@cbs.dtu.dk)

Alsmark, C., Frank, A., Karlberg, E. O. & 10 other authors (2004). The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc Natl Acad Sci U S A* **101**, 9716–9721.

Barbe, V., Vallenet, D., Fonknechten, N. & 12 other authors (2004). Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic Acids Res* (in press).

Bell, K. S., Sebaihia, M., Pritchard, L. & 29 other authors (2004). Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors. *Proc Natl Acad Sci U S A* Epub ahead of print, doi:10.1073/pnas.0402424101

Dujon, B., Sherman, D., Fischer, G. & 64 other authors (2004). Genome evolution in yeasts. *Nature* **430**, 35–44.

Pedersen, A. G., Jensen, L. J., Brunak, S., Staerfeldt, H. H. & Ussery, D. W. (2000). A DNA structural atlas for *Escherichia coli*. *J Mol Biol* **16**, 907–930.

Schnepf, E., Crickmore, N., Van Rie, J., Lereclus, D., Baum, J., Feitelson, J., Zeigler, D. R. & Dean, D. H. (1998). *Bacillus thuringiensis* and its pesticidal crystal proteins. *Microbiol Mol Biol Rev* **62**, 775–806.

Ussery, D. W. & Hallin, P. F. (2004). Genome Update: AT content in sequenced prokaryotic genomes. *Microbiology* **150**, 749–752.

DOI 10.1099/mic.0.27504-0